VISEM: A Multimodal Video Dataset of Human Spermatozoa

Trine B. Haugen
Oslo Metropolitan University, Norway
Faculty of Health Sciences, Norway

Steven A. Hicks
SimulaMet, Norway
Oslo Metropolitan University, Norway

Jorunn M. Andersen
Oslo Metropolitan University, Norway
Faculty of Health Sciences, Norway

Oliwia Witczak
Oslo Metropolitan University, Norway
Faculty of Health Sciences, Norway

Hugo L. Hammer
Oslo Metropolitan University, Norway

Rune Borgli
Simula Research Laboratory, Norway
University of Oslo, Norway

Pål Halvorsen
SimulaMet, Norway
Oslo Metropolitan University, Norway

Michael Riegler
SimulaMet, Norway

ABSTRACT

Real multimedia datasets that contain more than just images or text are rare. Even more so are open multimedia datasets in medicine. Often, clinically related datasets only consist of image or videos. In this paper, we present a dataset that is novel in two ways. Firstly, it is a multi-modal dataset containing different data sources such as videos, biological analysis data, and participant data. Secondly, it is the first dataset of that kind in the field of human reproduction. It consists of anonymized data from 85 different participants. We hope this dataset paper will inspire people to apply their knowledge in this important field, generate shareable results in the domain, and ultimately improve human infertility investigation and treatment.

CCS CONCEPTS

• **Applied computing** → **Health informatics**; • **Computing methodologies** → *Visual inspection*; *Neural networks*; *Classification and regression trees*;

KEYWORDS

Male fertility, Semen analysis, Spermatozoa, Machine Learning, Artificial Intelligence, Videos, Images, Dataset
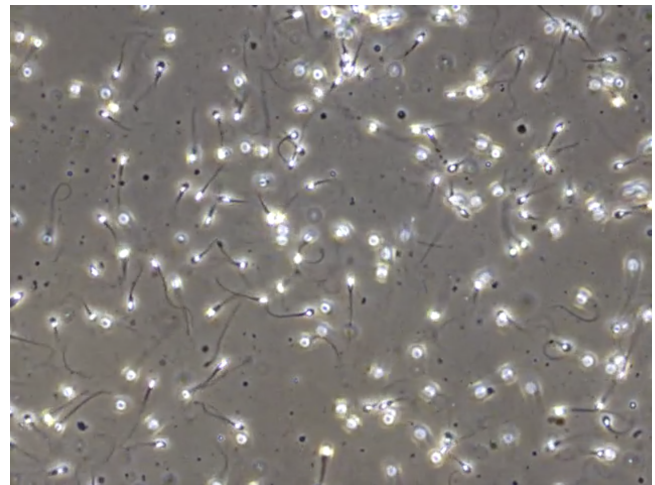
Figure 1: Example frame of a video in the dataset showing a typical semen sample. The bright dots are the sperm heads, and the black lines are the tails.

1 INTRODUCTION

In the field of multimedia research, a trend towards medical applications can be observed. This fact is underlined by a "Brave New Topic" proposal of multimedia in medicine [26], multiple medical multimedia tutorials [25, 27], and several emerging medical related workshops and special sessions in multimedia related conferences[1,2]. However, multi-modal clinical datasets are still rare, as most datasets focus on images with little to no associated attributes. In this paper, we present a dataset containing more than just images and videos, but also clinical attributes which may be combined to learn hidden relationships between the different modalities [3].

In the domain of medical multimedia, we are currently looking at the area of human reproduction by analyzing microscopic images of sperm as shown in Figure 1. During the last few decades, fertility rates in many industrialized countries have on average fallen far below 2.1, which is considered to be the threshold to sustain a population level [22]. In Norway, fecundity has decreased from 1.98 in 2009 to 1.62 in 2017 [21]. This decline can partly be explained by socioeconomic factors, but it is also due to biologic matters. One in six couples will encounter infertility [4], with 40 percent being accounted to both male and female infertility factors. 10-15

[1] http://www.multimediaeval.org/mediaeval2018/medico/
[2] https://mmhealth.uni-oldenburg.de

percent of the cases go unexplained, based on the current methods for evaluating fecundity.

Several studies have indicated that sperm count has declined globally during the last decades [6, 18]. Furthermore, geographical differences in semen quality have been observed. Among conscripts (participants in a study) from the Nordic and Baltic countries, the poorest semen quality was seen in Denmark and Norway [17]. Semen quality is associated with fertility, but the results are inconsistent with regards to which parameter is the best predictor [5, 15, 30]. Furthermore, the function of the spermatozoon (a mature and motile male gamete) is not necessarily reflected by the traditional semen parameters.

Semen analysis is often performed early in the investigation of involuntary childlessness. The clinical value of the semen analysis is, however, debated. The results are not easily interpreted and must be related to the participant history and other examinations. The World Health Organization (WHO) has developed guidelines for examination of human semen, first published in 1980 and updated four times, with the last edition published in 2010 [23]. Still, the assessments are based on a methodology introduced in the 1950s. Standardization is important to achieve accurate and reproducible results and to obtain evidence-based reference values for semen characteristics. However, assessments often vary between different laboratories, mainly because they are based on subjective evaluations. The standard semen analysis recommended by WHO includes evaluation of sperm concentration, total sperm count, sperm motility, sperm morphology, and sperm vitality. In some cases, additional examinations are performed, like testing for antibody coating of spermatozoa and assessment of leukocytes in semen.

During the 1980s, the introduction of personal computers and improved digitalization of images made it possible to develop computer-aided sperm analysis (CASA). The first commercial CASA instrument appeared in 1985. The hope was that CASA would contribute to a more rapid and objective assessment of the spermatozoa, but due to the nature of semen samples, it has been difficult to obtain accurate and reproducible results. Thus, CASA has not been recommended for clinical use. However, during the last years, the technology has improved and may work well with prepared semen samples, i.e., spermatozoa isolated from seminal plasma and suspended in a medium. An advantage with CASA is that the sperm kinematic may be recorded, not only the categories progressive motile, non-progressive motile, and immotile spermatozoa. It is still not well elucidated what clinical value the motility pattern provides, but the improvement of methods makes it possible to obtain more reliable data. Imaging of sperm movements may, however, provide information that is not captured by the CASA systems either. By using machine learning (ML) methods for visual recognition, such patterns may be revealed.

There may be other characteristics of the spermatozoa that provide information of relevance to fertility. These may be molecular markers or constituents of the spermatozoa. The membrane of the spermatozoon is especially rich in long unsaturated fatty acids, and in a previous study, we have shown that the sperm fatty acid composition is associated with semen quality [3]. Furthermore, we found that being overweight is negatively related to semen quality, and also that the fatty acid profile of the sperm of obese men may be

unfavorable. It is expected that data collected during the study period has a potential for automatic analysis, but also for novel ways to interact and search within the data. Possible research questions range from tracking sperms in real-time to automatically assess the quality of semen-based on videos. In addition, the video data could be combined with other data collected from the participants.

Sperm-related data faces the same problems as other clinical datasets. First of all, it is often difficult to share data due to legal requirements. Secondly, the knowledge about what the data contains and which are interesting medical research questions are often hard to find for researchers not familiar with the field. Finally, datasets are often also small compared to what would be effective for a proper analysis and evaluation of the results. We try to tackle the above-mentioned challenges by providing a clinical dataset in an open and explained way to experts not familiar with the medical field. The goal is to encourage these researchers to explore a new and exciting medical domain and contribute to the society with their research. In addition, we encourage comparable and open research also in the medical field where data access is usually difficult and often restricted to a limited number of researchers. The main contributions of this paper therefore are:

(1) Presentation of and making available a novel, multi-modal, open dataset, from a large number of participants, containing videos, semen characteristics, and demographic data, all anonymized.
(2) A baseline evaluation of the two most obvious research questions within the dataset and suggestions for additional future research questions.

To the best of our knowledge, the dataset contains more samples and far more attributes per participant than any sperm dataset openly available today. The dataset opens up for a wide range of new and interesting analyses, and a proper and fair comparison between different methods, both from a medical and a multimedia perspective. In the following, the process of collecting the data and the data itself are described. Moreover, a baseline evaluation is presented, including suggestions for future research directions using the present dataset.

2 DATA COLLECTION

The presented data was originally collected for studies on how overweight and obesity relate to the male reproductive function [3, 4]. Participants in the study were males aged 18 years or older and were recruited between 2008 and 2013 from the normal population through advertisements in newspapers and weight loss programs, and patients from obesity and fertility clinics. Further details on the recruitment have been described previously by Andersen et al. [2]. The study was approved by the Regional Committee for Medical and Health Research Ethics, South East, Norway, and all participants provided written informed consent. The project was closed in December of 2017, and all data was anonymized.

Participants in the study provided semen samples, and standard semen analysis was performed according to WHO recommendation [23], including assessment of sperm motility, sperm concentration and total sperm count, ejaculate volume, sperm morphology, and sperm vitality. For video recording for sperm motility, a sample was placed on a heated microscope stage (37℃) and examined

under a 400× magnification using an Olympus CX31 microscope. Videos were captured by a microscope mounted camera (specifically a UEye UI-2210C made by IDS Imaging Development Systems in Germany) and saved as an AVI file. Motility analysis was performed based on the videos. Fatty acids from spermatozoa and serum phospholipids were extracted from the samples and analyzed by gas chromatography as described in [3]. Sex hormones were measured in blood samples as described in [2]. Additionally, AMH was measured in seminal plasma. For the association studies, multiple linear regression and partial correlation were used for statistical analyses, with *SPSS Statistics 20* as the analysis tool.

2.1 Medical background

In some studies, sperm motility has been shown to be associated with fertility [16, 20], whereas the predictive value has been shown to be limited in other studies [5, 30]. The sperm motility is assessed by microscoping of a fresh, liquefied semen sample. Spermatozoa in the various motility categories are counted; progressive, non-progressive, and immotile. Progressive motility is defined as spermatozoa moving linearly or in a large circle, regardless of speed, non-progressive motility as all other patterns of motility with an absence of progression, and immotility as no movement. 200 spermatozoa in each of the two replicates are counted, and the average percentage is calculated.

The evaluation of sperm morphology is controversial, but although the WHO 2010 manual recommends the use of strict criteria [23], different methodologies may have prevented a consensus across laboratories. To assess the morphology, fixed and stained semen is prepared and examined with brightfield optics at 400× to 1000× magnification with oil immersion. Approximately 200 spermatozoa per replicate are assessed for the percentage of normal and abnormal forms as well as the various defects; head, tail, midpiece, and the presence of cytoplasmic droplets. Reference ranges for the semen variables have been established based on fertile men, who are defined as men whose partner conceived within 12 months after stopping the use of contraception [10]. A one-sided reference interval has been considered to be more appropriate for semen parameters since high values are unlikely to be detrimental to fertility. The methodological approach with ML is intended to gain more information from a semen sample than obtained by the standard WHO analysis and a CASA system.

2.2 Related work

Open datasets in sperm analysis are rare and usually also very limited in terms of size (number of images, videos, etc.) and participant diversity (number of different participants included) Another problem is usually that participants are recruited from the clinics meaning they are not representative for the general population. This makes it hard to evaluate algorithms properly, but also to compare different methods. Currently, only two small image datasets exist. The first dataset, called Human Sperm Head Morphology dataset (HuSHeM) [29], contains images of semen samples from 15 participants. In total, 725 images were taken during the study, but only 216 are openly available. The images have a resolution of 131×131 pixels and are classified into four different classes (normal, tapered, pyriform, and amorphous). For each of the images, a binary mask of the sperm in the image is provided. The second image

dataset is the SCIAN-MorphoSpermGS dataset [7]. It contains a total of 20 images of sperm samples. In addition to images, it also contains binary maps of the spermatozoa shown in the images. The maps are separated into three categories depicting segments of parts of the sperms (acrosome masks, head masks, and nucleus masks). In addition to these two datasets, Gil et al. [14] present a dataset containing semen analysis of 100 volunteers. The dataset contains only features from the analysis, and no images or videos are provided.

In terms of methods used to analyze sperms, several related works can be found. All related works use datasets not made public beside of the three small datasets presented above. Morphological classification of sperm can either be applied to living and moving sperms [13] or based on images of fixed and stained sperms. For the second category, a set of image features is selected which is used to train statistical or ML classifiers. The features can either be selected manually based on quantities like sperm head length, width and area [8, 24, 28], or automatically. Within automatic feature selection, Yi et al. [32] used a discrete Fourier transform and a dyadic wavelet transformation to characterize sperm head boundaries. Li et al. [19] extracts image features using principal component analysis and classifies using k-nearest neighbors. This approach outperformed selecting features based on scale-invariant feature transform or classifying based on a back propagating neural network. Shaker et al. [29] used squared patches extracted from sperm head images train a dictionary learning model.

In summary, current analyses of semen quality show potentials to develop improved tools for fertility investigation and sperm selection in particular. However, with more information, the potential can be even higher. New datasets containing information beyond images and videos only are therefore required.

3 DATASET DETAILS

VISEM contains data from 85 male participants aged 18 years or older. For each of the participants, parameters from a standard semen analysis, a video of live spermatozoa, sperm fatty acid profile, the fatty acid composition of phospholipids of serum, demographic data and WHO analysis data are available. For some participants, two video files were made since there was a drift in the first sample recorded. This makes it difficult for the laboratory personnel to assess the motility. We decided to only include one video per participant due to dataset size concerns. The dataset contains over 35 gigabytes of videos, with each video lasting between two to seven minutes. The resolution of the videos is 640 × 480, and the frame-rate is 50 frames per second.

The dataset contains in total six CSV-files (five for data and one for the video to participant ID mapping), a description file, and a video folder and can be accessed via **https://datasets.simula.no/visem** or **https://github.com/simula/mmsys2019-VISEM**. Each of the video files is named with an ID, the date of video capture and a small optional description. Then, the end of the filename contains the code of the person who assessed the video using the WHO standard. In Figure 1, a sample frame from one of the videos is shown. In this specific sample, which is defined as a normal sample, a lot of spermatozoa can be seen. The distance from the microscope to the sample is fixed. Furthermore, the camera is kept in a fixed

position. This could also make it possible for example to estimate size differences or even try to perform a 3D reconstruction.

Further, VISEM contains five CSV-files for each of the other data provided, a CSV-file with the IDs linked to each their video, and a text file containing descriptions of some of the columns of the CSV-files. One row in each CSV-file represents a participant. The provided CSV-files are:

- **semen_analysis_data**: The results of standard semen analysis.
- **fatty_acids_spermatozoa**: The levels of several fatty acids in the spermatozoa of the participants.
- **fatty_acids_serum**: The serum levels of the fatty acids of the phospholipids (measured from the blood of the participant).
- **sex_hormones**: The serum levels of sex hormones measured in blood of the participants.
- **participant_related_data**: General information about the participants such as age, abstinence time and Body Mass Index (BMI).
- **videos**: Overview of which video-file belongs to what participant.

The dataset can be used for research without any restrictions. Commercial use needs to be approved by the authors.

4 APPLICATIONS OF THE DATASET

We hope that this dataset may help researchers develop new methods for automatically detecting and predicting different aspects of human fertility. For example, predicting the motility and morphology of sperms would go a long way in reducing a doctor's workload. Motility and morphology are key attributes for determining the quality of a given sperm sample. Motility tells us something about the individual movement of each sperm, while morphology tells us something about the shape and form of the sperm cells. Another potential use-case is tracking individual sperms in real-time using the presented videos or perform semen quality analysis using the included fatty acid data. Furthermore, using the data collected from the WHO analysis, semen quality could be presented as a classification task. Some possible research questions which could be interesting to address using this dataset are (but not limited to):

- Is it possible to perform real-time tracking of spermatozoa in the videos. This could be very helpful for medical personnel to keep track of spermatozoa during the analysis.
- Is it possible to predict motility or morphology attributes from the videos only? This could save medical personnel a lot of time used to perform a manual analysis.
- Can a combination of different data sources improve the performance of prediction or tracking? This could be interesting in the sense of improving the performance in general but also to find new connections within the different modalities.
- How are different data sources related to semen quality? For example, are certain fatty acids related to semen quality and how? This could be very helpful to improve current knowledge and treatment but also to find completely new medical evidence.

In addition to these possible research questions, the dataset will hopefully inspire researchers to approach even more possible applications and research directions. Possible directions could be for example be in the fields of segmentation, video analysis, information retrieval, ML, object detection, computer vision, data fusion, and medical multimedia systems. Another important aspect is that the VISEM dataset also allows an easy and fair comparison between different methods. This is often a challenge in health-related data since the data is often restricted due to legal issues. Therefore, being one of the first open medical multi-modal datasets available to the multimedia community, we encourage researchers to explore the dataset with their methods and skills.

5 SUGGESTED METRICS

As pointed out in Section 4, the presented dataset contains several different research directions and questions to answer. First, the dataset may be used for typical regression tasks, such as predicting the motility and morphology of sperms for a given sample. Second, the dataset can be used for classification tasks using the data gathered through CASA analysis. Given this wide range of use-cases, it is hard to suggest good metrics for the dataset. Nevertheless, we recommend to at least use the most basic measures for each category. For classification, we suggest precision, recall (also called sensitivity in a medical context), accuracy, specificity, Matthews correlation coefficient, and F1 score. Metrics typically seen as most important in a medical context are sensitivity and specificity. If the data is used to perform predictions, e.g., one wants to predict the percentage of living sperms in a given sample based on a video, measures such as mean squared error (MSE) or mean absolute error (MAE) are advised. Additionally, we suggest using leave-one-out cross-validation to obtain more robust and generalized results. In any case, as one can see, the chosen metrics for this dataset strongly depend on the use-case and should be chosen accordingly.

6 BASELINE PERFORMANCE

In order to provide a baseline for future users of the dataset, we performed two different experiments. In both, the goal is to predict the quality of sperm samples in terms of motility and morphology. Motility is represented by the attributes percentage of progressive sperms (straight forward progression), percentage of non-progressive sperms (not forward progression/slow progression), and percentage of immobile sperms (no progression at all).

Morphology is defined by percentage of normal sperms (normality of the sperms), percentage of sperms with head failures (head defect), percentage of sperms with midpiece failures (midpiece defect), percentage of sperms with tail failures (tail defect), percentage of sperms with cytoplasmic droplet (cytoplasmic droplet is a defect), and the teratozoospermia index (defined as the number of abnormalities present per abnormal spermatozoon).

For experiment one, we used the videos as an input to predict the attributes for motility and morphology. For experiment two, we use the sex hormones and participant-related data as input. For both experiments, we perform three-fold cross-validation and report the MAE. The methods for the two experiments are explained in the following.

6.1 Experiment 1: Videos only

For the experiments using videos only to make predictions, we based our approach on deep learning algorithms, specifically deep convolutional neural networks (CNNs). Due to the computational complexity of using 3D CNNs, we limited ourselves to performing single frame prediction by extracting frames from the included videos and passing them through a 2D CNN to perform regression. For our implementation, we used Keras [9] with TensorFlow [1] as a back-end. The model used for prediction is based on an Inception V3 architecture [31], where we used the implementation included in Keras and trained it from scratch. We experimented using transfer learning from the ImageNet [11] weights included in Keras, but quickly found that it neither improved performance or time to convergence. Additionally, we performed the same experiments using other popular CNN architectures such as DenseNet and VGG, but have opted to only include the network which gave us the best results. As previously mentioned, the models trained for each experiment is done using three-fold cross-validation, where we extracted 250 frames, at evenly dispersed intervals, from each video (21,250 frames total) and used 14,250 for training and 7,000 for validation. All networks were trained using Nadam [12] as the optimizer and using MAE to calculate loss.

6.2 Experiment 2: Sex hormones and participant related data

For the experiments using sex hormones and participant-related data, we use three simple and well-known methods. As a baseline, we use a ZeroR regression (also called pseudo regression). The cross-validation coefficient for these methods is defined with Q2=0. Attributes predicted will be equal to the average calculated over the whole training set. The second method we are using is linear regression. The final method is Random Forest for regression which is an additive model using a sequence of base models to perform the final prediction. For each method, we trained one model per attribute.

6.3 Baseline results

Since we are only interested in the motility and morphology of sperm as a whole, and the mean absolute error is additive, we only report the overall error calculated by summing up the individual errors. Table 1 shows the prediction performance for both motility and morphology prediction of all proposed methods.

Starting with methods for predicting sperm motility, we see that all techniques outperform the ZeroR baseline, with the deep CNN approach achieving an overall best score of 31.6157 (a 7.0740 difference from the baseline). This indicates a good performance and shows that we are able to predict the motility and mobility of sperms based on single video frames alone. It is important to point out that even if the results are promising they are not applicable in a clinical setting with further improvement. Furthermore, looking at the methods using only sex hormones and participant data for prediction (early fusion with linear regression and random forest), we see that these methods too beat the baseline by quite a large margin (2.8324 for random forests and 1.4805 for linear regression). For predicting the sperm morphology, we see that two

Table 1: Prediction performance in terms of Mean Absolute Error.

Method	Predicted	Mean Absolute Error
Inception V3	Motility	31.6157
Linear regression	Motility	37.2092
Random Forest regression	Motility	35.8573
Baseline (ZeroR)	Motility	38.6897
Inception V3	Morphology	21.5697
Linear regression	Morphology	23.0028
Random Forest regression	Morphology	21.0614
Baseline (ZeroR)	Morphology	21.9202

of the three proposed methods are able to outperform the baseline, with random forest regression having the best performance at 0.8588 lower MAE than the baseline. The linear regression was not able to beat the baseline at 23.0028. Although the difference in error between methods is lower than that for motility, it is still an improvement and shows that these techniques work better than simply predicting the average. Nevertheless, this may indicate that sperm morphology is harder to predict based on just sex hormones and participant-related data, or static video frames alone. For all performed experiments, the optical flow between frames was not taken into account, which would most likely lead to further improve the results described.

7 CONCLUSION

Reproducibility and comparability of results are seen as the basis of good research. In this paper, we presented a dataset in a domain that usually comes with the challenge of data sharing. We presented a multi-modal dataset in the field of human reproduction collected from 85 participants. The dataset contains videos, analysis data, and participant-related data. In addition, we provide a baseline analysis and possible research questions that can be addressed with the dataset. We hope that the dataset will be useful for researchers in the field of human reproduction but also that it will encourage scientists not familiar with the topic to try their approaches and help to improve the outcome.

ACKNOWLEDGEMENTS

REFERENCES

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. (2015). https://www.tensorflow.org/ Software available from tensorflow.org.
[2] Jorunn M Andersen, Hilde Herning, Elin L Aschim, Jøran Hjelmesæth, Tom Mala, Hans Ivar Hanevik, Mona Bungum, Trine B Haugen, and Oliwia Witczak. 2015. Body mass index is associated with impaired semen characteristics and reduced levels of anti-müllerian hormone across a wide weight range. *PloS one* 10, 6 (2015), e0130210.

[3] Jorunn M Andersen, Per Ola Rønning, Hilde Herning, Siri D Bekken, Trine B Haugen, and Oliwia Witczak. 2016. Fatty acid composition of spermatozoa is associated with BMI and with semen quality. *Andrology* 4, 5 (2016), 857–865.

[4] Jacky Boivin, Laura Bunting, John A. Collins, and Karl G. Nygren. 2007. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. *Human Reproduction* 22, 6 (2007), 1506–1512.

[5] Jens Peter E Bonde, Erik Ernst, Tina Kold Jensen, Niels Henrik I Hjollund, Henrik Kolstad, Thomas Scheike, Aleksander Giwercman, Niels Erik Skakkebæk, Tine Brink Henriksen, and Jørn Olsen. 1998. Relation between semen quality and fertility: a population-based study of 430 first-pregnancy planners. *The Lancet* 352, 9135 (1998), 1172–1177.

[6] Elisabeth Carlsen, Aleksander Giwercman, Niels Keiding, and Niels E Skakkebæk. 1992. Evidence for decreasing quality of semen during past 50 years. *British Medical Journal* 305, 6854 (1992), 609–613.

[7] Violeta Chang, Alejandra Garcia, Nancy Hitschfeld, and Steffen Hrtel. 2017. Gold-standard for Computer-assisted Morphological Sperm Analysis. *Comput. Biol. Med.* 83, C (April 2017), 143–150. https://doi.org/10.1016/j.compbiomed.2017.03.004

[8] Violeta Chang, Laurent Heutte, Caroline Petitjean, Steffen Härtel, and Nancy Hitschfeld. 2017. Automatic classification of human sperm head morphology. *Computers in biology and medicine* 84 (2017), 205–216.

[9] François Chollet and others. 2015. Keras: Deep learning library for theano and tensorflow. *URL: https://keras. io/k* (2015).

[10] Trevor G Cooper, Elizabeth Noonan, Sigrid Von Eckardstein, Jacques Auger, HW Baker, Hermann M Behre, Trine B Haugen, Thinus Kruger, Christina Wang, Michael T Mbizvo, and others. 2010. World Health Organization reference values for human semen characteristics. *Human reproduction update* 16, 3 (2010), 231–245.

[11] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.

[12] Timothy Dozat. 2015. Incorporating Nesterov Momentum into.

[13] Fatemeh Ghasemian, Seyed Abolghasem Mirroshandel, Sara Monji-Azad, Mahnaz Azarnia, and Ziba Zahiri. 2015. An efficient method for automatic morphological abnormality detection from human sperm images. *Computer methods and programs in biomedicine* 122, 3 (2015), 409–420.

[14] David Gil, Jose Luis Girela, Joaquin De Juan, M Jose Gomez-Torres, and Magnus Johnsson. 2012. Predicting seminal quality with artificial intelligence methods. *Expert Systems with Applications* 39, 16 (2012), 12564–12573.

[15] David S Guzick, James W Overstreet, Pam Factor-Litvak, Charlene K Brazil, Steven T Nakajima, Christos Coutifaris, Sandra Ann Carson, Pauline Cisneros, Michael P Steinkampf, Joseph A Hill, and others. 2001. Sperm morphology, motility, and concentration in fertile and infertile men. *New England Journal of Medicine* 345, 19 (2001), 1388–1393.

[16] Trine B Haugen, Thore Egeland, and Øystein Magnus. 2006. Semen parameters in Norwegian fertile men. *Journal of andrology* 27, 1 (2006), 66–71.

[17] Niels Jørgensen, Elisabeth Carlsen, Ingrid Nermoen, Margus Punab, Jyrki Suominen, Anne-Grethe Andersen, Anna-Maria Andersson, Trine B Haugen, Antero Horte, Tina Kold Jensen, and others. 2002. East–West gradient in semen quality in the Nordic–Baltic area: a study of men from the general population in Denmark, Norway, Estonia and Finland. *Human Reproduction* 17, 8 (2002), 2199–2208.

[18] Hagai Levine, Niels Jørgensen, Anderson Martino-Andrade, Jaime Mendiola, Dan Weksler-Derri, Irina Mindlis, Rachel Pinotti, and Shanna H Swan. 2017. Temporal trends in sperm count: a systematic review and meta-regression analysis. *Human Reproduction Update* 23, 6 (2017), 646–659.

[19] Jiaqian Li, Kuo-Kun Tseng, Haiting Dong, Yifan Li, Ming Zhao, and Mingyue Ding. 2014. Human sperm health diagnosis with principal component analysis and K-nearest neighbor algorithm. In *Medical Biometrics, 2014 International Conference on*. IEEE, 108–113.

[20] Gunilla Malm, Lars Rylander, Aleksander Giwercman, and Trine B Haugen. 2018. Association between semen parameters and chance of fatherhood - a long-term follow-up study. *Andrology* (2018). https://doi.org/10.1111/andr.12558 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/andr.12558

[21] United Nations. 2017. United Nations population facts. http://www.un.org/en/development/desa/population/publications/pdf/popfacts/PopFacts_2017-3.pdf. (2017). Accessed: 2019-01-21.

[22] Statistics Norway. 2018. SSB birth per year in Norway. https://www.ssb.no/befolkning/statistikker/fodte/aar. (2018). Accessed: 2019-01-21.

[23] World Health Organization and others. 2010. WHO laboratory manual for the examination and processing of human semen. (2010).

[24] F Pérez-sánchez, JJ de Monserrat, and C Soler. 1994. Morphometric analysis of human sperm morphology. *International journal of andrology* 17, 5 (1994), 248–255.

[25] Michael Riegler, Pål Halvorsen, and Bernd Münzer nad Klaus Schöffman. 2018. The Importance of Medical Multimedia. In *Proceedings of the ACM International Conference on Multimedia (MM)*. 2106–2108.

[26] Michael Riegler, Mathias Lux, Carsten Griwodz, Concetto Spampinato, Thomas de Lange, Sigrun L Eskeland, Konstantin Pogorelov, Wallapak Tavanapong, Peter T Schmidt, Cathal Gurrin, Dag Johansen, Håvard Johansen, and Pål Halvorsen. 2016. Multimedia and Medicine: Teammates for Better Disease Detection and Survival. In *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, 968–977.

[27] Klaus Schöffmann, Bernd Münzer, Michael Riegler, and Pål Halvorsen. 2017. Medical Multimedia Information Systems (MMIS). In *Proceedings of the ACM International Conference on Multimedia (MM)*. 1957–1968.

[28] Fariba Shaker, S Amirhassan Monadjemi, and Javad Alirezaie. 2017. Classification of human sperm heads using elliptic features and LDA. In *3rd International Conference on Pattern Recognition and Image Analysis (IPRIA)*. IEEE, 151–155.

[29] Fariba Shaker, S Amirhassan Monadjemi, Javad Alirezaie, and Ahmad Reza Naghsh-Nilchi. 2017. A dictionary learning approach for human sperm heads classification. *Computers in biology and medicine* 91 (2017), 181–190.

[30] Remy Slama, Florence Eustache, B Ducot, Tina Kold Jensen, Niels Jørgensen, Antero Horte, S Irvine, J Suominen, AG Andersen, J Auger, and others. 2002. Time to pregnancy and semen parameters: a cross-sectional study among fertile couples from four European cities. *Human Reproduction* 17, 2 (2002), 503–515.

[31] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. 2015. Rethinking the Inception Architecture for Computer Vision. *CoRR* abs/1512.00567 (2015). arXiv:1512.00567 http://arxiv.org/abs/1512.00567

[32] WJ Yi, KS Park, and JS Paick. 1998. Parameterized characterization of elliptic sperm heads using Fourier representation and wavelet transform. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, Vol. 2. IEEE, 974–977.